



Computing and Information Technology Student Mini Showcase

Abstracts LA-UR-12-23517

PlateSpin PowerRecon- Planning for VM Growth **Marisa Boettcher — University of New Mexico - Los Alamos** *Mentor: Moses Gallegos* Group: NIE-IS

With the program PlateSpin PowerRecon we are able to monitor, plan, and use its management features for providing information about NIE-IS owned servers in the data center. This is useful because the program can help decide which servers can be consolidated and can handle more of a workload. As an example, if three different servers have alternating peak usage times, these assets can most likely be combined. Thus reducing assets used, power consumption, and cooling costs. PlateSpin PowerRecon can also be helpful in the decision process of what servers would be good candidates for transfer to virtualization based on CPU, disk drives, applications, and operating system.

GlusterFS: One Storage Server to Rule Them All **Eric Boyer — Michigan Technological University** **Matthew Broomfield — New Mexico Institute of Mining and Technology** **Terrell Perrotti — South Carolina State University** *Mentors: David Kennel and Greg Lee, DCS-1* Group: INST-OFF

The need for a unified storage system that is fast, resilient, and economically viable is constantly increasing. In this project we evaluate the usability of the GlusterFS storage system as both a high performance storage system and a general purpose storage system. We assess the performance of various GlusterFS configurations, along with its capability to support common enterprise storage features. Finally, we analyze the overall ease of management for a GlusterFS system.

Building a Parallel Cloud Storage System using OpenStack's Swift Object Store and Transformative Parallel I/O

Andrew Burns — New Mexico Institute of Mining and Technology
Kaleb Lora — University of Washington – Seattle
Esteban Martinez — University of New Mexico
Martel Shorter — Prairie View A&M University
Mentors: HB Chen, Parks Fields, and Benjamin McClelland, HPC-5;
David Sherrill and Pamela Smith, HPC-3; Alfred Torrez, HPC-1
Group: INST-OFF, HPC-5

In this project we built a prototype of a parallel cloud storage system. We started by determining the scaling capability of Swift's object storage system coupled with the parallel I/O feature from the LANL's Parallel Log-based File System (PLFS). PLFS is used to parallelize and store data in an N-1-N fashion. We wanted to see if and how PLFS would be incorporated in a cloud environment.

Swift is an open source cloud storage application used for creating redundant, scalable object storage using clusters of standardized servers. Swift can store petabytes of accessible data, and it serves as a long-term storage system for more permanent, static data. That data can be retrieved, leveraged, and then updated if necessary.

We then applied S3QL on top of Swift. S3QL is an active Python/FUSE-based file system that runs with Amazon's "Simple Storage Service (S3)". S3 can be transposed into a full-featured UNIX file system, which is usable by PLFS. This then serves as the file system interface between LANL's PLFS and OpenStack's Swift object store.

We interacted with PLFS, Swift, and S3QL via bash and python. After successfully integrating the PLFS transformative parallel I/O feature with the Swift, we conducted various studies. Typically, we focused on the I/O bandwidth and system performance.

Stewardship of Resources in a Virtual Infrastructure

David Carlson — Lewis University

Mentor: Jeffery Click

Group: NIE-IS

Infrastructure on Demand also known as IoD is a lab wide virtual private cloud service currently hosting more than 600 virtual servers spanning many groups and divisions at LANL. IoD is a virtual infrastructure running on physical hardware comprised of HP C7000 enclosures with blade modules, each module contains 2 CPUs consisting of 12 Cores, and 144 GB of Memory. Layered on top of the physical hardware is the VMware ESXI hypervisor and the corresponding VMware vSphere suite of applications: vCenter, vCloud Director, View, vCenter Operations, etc.

As the Government is currently in a state of tightened budgets and LANL is looking to reduce costs; stewardship of resources has become a priority. Within a virtual infrastructure stewardship of resources is equally important, for many reasons. "Right sizing" a virtual machine is essential due to the nature of the environment in which it lives, an environment of shared resources. Were a virtual machine to be improperly sized it can adversely affect other machines within the infrastructure reducing the performance of other VMs along with itself. Improperly sized machines can also lead to the waste of physical resources and thus the need to prematurely purchase more physical hardware to host virtual machines on. This is where VMware Operations Manager comes in; it provides resource management for a VMware virtual infrastructure. Ops Manager can generate reports listing over-sized virtual machines, under-sized virtual machines, and capacity management to name the most beneficial reports. These reports can help an administrator ensure that an infrastructure is properly sized and thus ensure stewardship of resources.

DHCP

Alexander Castillo — New Mexico State University

Mentor: Brenda Joyce

Group: SAE-2

The project's purpose was to enable the use of Dynamic Host Configuration Protocol on the yellow network. All internet protocol addresses at the lab are assigned to every machine. Previously, users were forced to manually enter the IP they were assigned or they were denied network connectivity. With the success of the DHCP project, users should now be able to merely plug their computer into a working Ethernet port and their IP will be assigned automatically. This project required the writing of a program to organize IP addresses based on subnets and for Hostmaster to be correctly updated with current machines in order for the addresses to be assigned correctly.

Scalable Node Monitoring

Alexander Drotar — University of Colorado - Boulder

Erin Quinn — Fairleigh Dickinson University

Landon Sutherland — New Mexico State University

Mentors: Mike Mason and Jon Bringham, HPC-3

Group: INST-OFF

In the field of cluster computing, node monitoring is crucial. With an abundance of nodes, it is important to make sure that all of the nodes and their applications are running properly to maximize cluster efficiency. Without a node monitoring system in place on a cluster, failed applications can go unnoticed, which can drastically affect performance. This project involved designing and implementing a scalable, parallel tool for monitoring nodes on computer clusters using Component Based Tool Framework (CBTF). This monitoring tool uses Lightweight Data Metric Service (LDMS) to gather node information and using CBTF's inherent scalability, the new tool can operate on a cluster made up of any number of nodes. In this project, the tool was tested and analyzed on an 8-node cluster running Centos 6.2.

Windows 7 Desktop Infrastructure Standardization and System Administration for X-Division

Angelica S. Gallegos — University of New Mexico

Mentors: Erin Powers and Joseph Pana

Group: DCS-2

The XCS Team is anticipating the future requirement to retire or upgrade all systems running the Windows XP operating system. Los Alamos Site Office does not officially have a date for elimination of Windows XP on LANL networks, but LANL has set a target date of October 2013 for retirement of these systems from the yellow network. The XCS Team gave themselves an internal date of October 2012.

The goal of my project is to help the XCS Team eliminate the remaining Windows XP systems within X-Division by either upgrading current systems or replacing legacy systems with new systems running Windows 7 before the October 2012 date.

Reasons to upgrade from Windows XP to Windows 7 include: standardization of Windows platforms, standardization of 3rd party software, improved system security, reduced difficulties identifying and troubleshooting Windows related issues, and decrease time and expenditures spent on system administration and support.

Metadata Reduction in an Exascale File System

Jun He — Illinois Institute of Technology

Mentors: Mike Lang and John Bent, HPC-5; and Aaron Torres, HPC-3

Group: HPC-5/New Mexico Consortium

File system metadata is becoming larger and more complex as we move towards exascale computing. PLFS (Parallel Log-structured File System) is a virtual file system that has shown significant improvements over current parallel file systems. However, it suffers from slower reads and oversized memory footprints when its metadata, an index that maps between logical and physical offsets of a file, becomes very large. Fortunately, most applications present I/O patterns that can be taken advantage of in order to compress indices. We have implemented I/O access pattern recognition in PLFS and managed to reduce metadata size significantly if patterns do exist. Experimental results show that index sizes are reduced by several orders of magnitude in some cases and

read bandwidth is raised significantly, while the write performance is left unaffected. We also see potential for this approach to be extended to other I/O workloads and we plan on exploring usage of these patterns elsewhere.

Implementing New HPC Accounts Software

Daniel Illescas and Lucia Short, University of New Mexico

Mentors: Ben Santos, and Laura Davey

Group: HPC-3

HPC Accounts is software that primarily handles user management of accounts on LANL HPC clusters and support machines. It automates user account requests and helps project managers maintain accounts. HPC needed a newer version of the Accounts software to offer more functionality and robust experience to users, approvers, and administrators. We were tasked with adding new features, fixing existing bugs, and offering greater usability, all while presenting a more modern design. To accomplish this, we modified the existing code base that utilized the rapid development framework CakePHP. The new product should aid both support personnel as well as customers in all account management aspects.

Applied Parallel Metadata Indexing

Michael Jacobi — University of New Mexico

Mentors: Cody Scott and Aaron Torres

Group: HPC-3

The GPFS Archive is a parallel archive used by hundreds of users in the Turquoise collaboration network. It houses 4+ petabytes of data in more than 170 million files. Currently, users must navigate the file system to retrieve their data, requiring them to remember file paths and names. A better solution might allow users to tag data with meaningful labels and search the archive using standard and user-defined metadata, while maintaining security. Last summer, I developed the backend to a tool that adheres to these design goals. The backend works by importing GPFS metadata into a MongoDB cluster, which is then indexed on each attribute. This summer, I implemented security and developed the user interface for the search tool. To meet security requirements, each database table is associated with a single user, which only stores records that the user may read, and requires a set of credentials to access. The interface to the search tool is implemented using FUSE (Filesystem in Userspace). FUSE is an intermediate layer that intercepts file system calls and allows the developer to redefine how those calls behave. In the case of this tool, FUSE interfaces with MongoDB to issue queries and populate output. A FUSE implementation is desirable because it allows users to interact with the search tool using commands they are already familiar with. These security and interface additions are essential for a usable product.

Verification and Validation Assessments Under Risk Management Framework

Travis Johnson — California State University – San Bernardino

Mentor: Ray Tyler

Group: CA-1

With the growth of severe cyber threats in the United States, cyber security policies, processes and procedures must have the ability to adapt to meet these growing challenges. This year LANL introduced the Risk Management Framework (RMF) which initiated the beginning of a new approach to manage the lab's cyber security efforts. The basis of the RMF is the focus on cyber security risk rather than a compliance-based "check-the-box" model. Within the RMF the focus relies on maintaining acceptable residual risk levels for the institution through a robust continuous monitoring program. An element of this is the Verification and Validation (V&V) Assessments. The V&V assessment procedures utilize automated continuous monitoring tools, verifying appropriate mitigations are in place on system, and facilitates in the cyber security awareness. The RMF and

V&V process is essential in helping the Laboratory prepare for increased cyber threats and challenges throughout the LANL Information Technology infrastructure.

iSSH vs. Auditd: Intrusion Detection in High Performance Computing

David Karns — New Mexico State University

Katy Protin — University of North Carolina at Chapel Hill

Justin Wolf — California State University

Mentors: Graham Van Heule, and Jim Williams, HPC-3; Alex Malin, HPC-DO

Group: INST-OFF

Our goal in this project was to provide insight into intrusions in high performance computing, focusing on tracking intruders' motions through the system. Currently, pattern-matching tools are used to detect suspicious behavior, but these tools do not provide methods to track a hacker's motions once inside the system. We tested two tools, instrumented secure shell (iSSH) and the Linux Auditing Framework (Auditd) to see if they provided insight into whether a users behavior is malicious. We wanted to explore how each tool is implemented, which is more effective, and how they affect computer performance and cost. While doing this project, we worked to modify these tools so that they would catch more types of suspicious behavior. We tested each tool by attacking our computer cluster, then modifying them to catch each type of attack. In this project, we found that both tools have both limitations and strengths. iSSH is most useful for tracking keystrokes and reporting suspicious commands, but as a newer tool, it is not as well documented and was difficult to set up. Auditd is good for keeping daily logs of activities that can be (but are not necessarily) malicious, but does not have as many reporting capabilities as iSSH.

The Design and Implementation of a Multi-level Content-Addressable Checkpoint File System

Abhishek Kulkarni — Indiana University

Mentor: Michael Lang

Group: HPC-5/New Mexico Consortium

Long-running HPC applications guard against node failures by writing checkpoints to parallel file systems. Writing these checkpoints with petascale class machines has proven difficult and the increased concurrency demands of exascale computing will exacerbate this problem. To meet checkpointing demands and sustain application perceived throughput at exascale, multi-tiered hierarchical storage architectures involving solid-state burst buffers are being considered. In this paper, we describe the design and implementation of cento, a multi-level, content-addressable checkpoint file system for large-scale HPC systems. Cento achieves in-flight checkpoint data reduction across all compute nodes through compression and elimination of duplicate blocks over a series of checkpoints. Through a detailed analysis of checkpoint dumps, we assess the benefits of data reduction for scientific applications that are representative of production workloads. We observe up to 40% data reduction within a limited sample of representative workloads. Finally, experiments on existing systems show a decrease in checkpoint commit latencies by 5 to 20% reducing the load on the parallel file system.

Exploring the Design Tradeoffs for Exascale System Services through Simulation

Abhishek Kulkarni — Indiana University

Mentor: Michael Lang

Group: HPC-5/New Mexico Consortium

Technology trends indicate that exascale systems will comprise hundreds of millions to billions of heterogeneous cores, and each core will have limited memory capacity and bandwidth. Current HPC system designs focus a great deal of effort on optimizing the use of interconnection networks and storage, but suffer from the lack of scalable solutions for infrastructure and management. Owing to the extreme scale and higher rate of component

failures, system software and services will need to be failure-resistant, adaptive and self-healing so that sustained operation of the system is achieved without frequent manual intervention. Existing system services are still designed around a centralized server paradigm with at most a single fail-over server and hence are susceptible to single points of failure.

We deconstruct services into their most basic components and provide general service taxonomy to classify existing system services. We explore the services landscape through several service architectures and identify the design tradeoffs for scalable system services. We further simulate these service architectures at exascale, with potentially millions of clients handled by tens of thousands of servers. We estimate basic parameters such as memory consumption and communication complexity analytically and perform large-scale discrete event simulations to estimate complex parameters such as client-perceived throughput and overall service efficiency. Through simulation, we finally demonstrate how churn (servers joining and leaving the system) affects the performance and efficiency of different service architectures including centralized, distributed with global view, distributed with partial view.

Hard Disk/Solid State Drive Synergy in Support of Data-Intensive Computing

Ke Liu — Wayne State University

Mentor: Kei Davis

Group: CCS-7

Data-intensive applications are becoming increasingly common in high-performance computing. Examples include combustion simulation, human genome analysis, and satellite image processing. Efficient access of data sets is critical to the performance of these applications. Because of the size of the data today's economically feasible approach is to store the data files on an array of hard disks or data servers equipped with hard disks and managed by a parallel file system such as PVFS or Lustre wherein the data is striped over a (large) number of disks for high aggregate I/O throughout.

With file striping, a request for a segment of logically contiguous file space is decomposed into multiple sub-requests, each to a different server. While the data unit for this striping is usually reasonably large to benefit disk efficiency, the first and/or last sub-requests can be much smaller than the striping unit if the request does not align with the striping pattern, severely compromising hard disk efficiency and thus application performance.

We propose to exploit solid state drives (SSD), whose efficiency is much less sensitive to small random accesses, to enable the alignment of requests to disk with the data striping pattern. In this scheme hard disks mainly serve large, aligned, sequential requests, with SSDs serving small or unaligned requests, thus respecting the relative cost, performance, and durability characteristics of the two media, and thereby achieving synergy in performance/cost. We will describe the design of the proposed scheme, its implementation on CCS-7's Darwin cluster, and performance results.

My Summer Internship in a Nutshell

Joshua McKown — New Mexico Institute of Mining and Technology

Mentor: Brenna Lee

Group: NIE-CS

During my summer at LANL, my first project was to create a digital Wireless Certification form that took user input, verified it, logged it, and sent an email to the appropriate people. It was later expanded to include the certification forms for SSL Web Server Certificates, Local and Commercial. The local web server page takes the user information and verifies emails it to the appropriate person, but cannot verify anything on the red network. The commercial web server form takes the user information and checks it. It then sends a verification email to the

appropriate person, and then gives them a link to the Entrust website with instructions on how to proceed.

The second project that I have done is automating the secure CryptoCard request form. First, my new online form makes sure that the person logged in is a manager, and then it takes the user input, shows the user the request was submitted, and an email is sent to the appropriate people. We then expanded to include the CryptoCard signing page which logs user acceptance of cryptocards for security audits and archive purposes.

When I first started this job, my only experience in network programming was a summer class that taught the basics of HTML and Javascript. Since then, I have learned how to use PHP, expanded my knowledge of HTML, and learned the basics of CSS. As such, my first project was very time demanding, as I was learning another programming language, and a lot of the server specific commands.

My experience at LANL has been wonderful. I have worked with the Networking and Infrastructure Engineering - Core Services group (NIE-CS) to make computer forms, making the applicant and the approvers jobs much easier.

Determining Power Utilization Efficiency of a Segmented Data Center

Alynn Montoya-Wiüff – New Mexico State University

Edwin Williams – Arizona State University

Mentor: Ron Velarde

Group: HPC-2

Power and environmental monitoring are essential pieces of improving and keeping a data center running efficiently. In order to gain a comprehensive understanding of a data center's energy consumption, access to data must be available in order to continually refine daily operations in search of best practice methods per Department of Energy mandate.

The traditional methods for calculating power efficiencies cannot be used for segmented data centers, therefore a new approach must be formulated. In order to determine an accurate power utilization efficiency (PUE) there must be a way to distinguish between data center cooling, comfort cooling, and other remote building cooling. A ratio was established by using overall chiller flow, computer room air conditioner (CRAC) unit flow, and their associated temperature differences to verify the total cooling load dedicated to the data center. A critical factor of PUE is the total IT load which can be calculated by pulling meter information from the dedicated power distribution units (PDUs).

Applying new methods of measuring efficiencies in older data centers can prove to be a nightmare if the existing equipment does not have the capability to provide (if at all) accurate input and output values. Together the building automation and the Environet systems at the Laboratory Data Communication Center (LDCC) provide the information necessary to compute PUE.

Extracting Structure from Nebulous Objects

David Nilosek – Rochester Institute of Technology

Mentor: Amy Galbraith

Group: ISR-3

In recent years, due to the increase of computational power and speed, many advancements have been made in the automation of photogrammetry using sophisticated structure from motion (SfM) techniques taken from the field of computer vision. This work looks at using modern SfM techniques for feature extraction and reconstruction that provides a robust way to detect and reconstruct ill-defined features. This work uses two modern SfM work

flows: Bundler and Patch-Based Multi-View Stereo (PMVS). This work also looks at two types of nebulous features: those due to the imaging process and those due to diffuse structures within the imagery. Mid-wave infrared (MWIR) imagery is used for nebulous features caused by the imaging process. A video from YouTube of the plume from the start of the recent Waldo Canyon Fire is used for the diffuse structure target.

Proposed PurchaseIT Website Revisions
Cameron Ranken — University of New Mexico
Mentor: Mike Mikus
Group: DCS-3

This summer I have been involved with updating and upgrading the Central Distribution Center's website, PurchaseIT.lanl.gov. Over the past two months, I have gained extensive knowledge of the laboratory's processes and exceptions involving the procurement of new information technology hardware. Through hands on experience with new IT hardware entering the lab, along with research, I have acquired a detailed understanding of the steps involved in procuring standard and non-standard information architecture. Included on this poster are the proposed revisions for several of the CDC's current web pages dedicated to the explanation of LANL's IT procurement policies, along with additional new documentation to be added to the website.

The Importance of Portable Document Format Files and Digital Signatures
Lucas Sandoval — New Mexico Institute of Mining and Technology
Mentors: John McDermion and Stacy Baker
Group: DCS-2

A Portable Document Format, or better known as a PDF, is a formal open standard known as ISO 32000. PDF files are viewable and printable on virtually any platform ranging from Windows and Mac OS to mobile platforms such as Android. PDF files provide easy access and availability to online data. Unlike other document viewing programs, such as Adobe PostScript, PDF files can be accessed at will. PDF files are important due to many key factors. PDF files are well known for their capability to incorporate hyperlinks, bookmarks, text, graphics and multimedia content. Since PDF files are able to store mass amounts of information without using mass amounts of computer memory, they have been able to change the way consumers view information on their PC. A consumer also has the capability of easily signing a document with a digital signature. This prevents hackers or other thieves from altering and stealing private information that is in a PDF file. As a student, I will be researching and pursuing the viability of maintaining document databases on servers using the Adobe document creation, signing, and tracking software to reduce the amount of paper used and stored and to enable greater accessibility to these documents for group management and AD/DO requests.

SSH Trust Relationship Analyzer
Marc Santoro — California State University, San Bernardino Campus
Mentor: Jim Gore
Group: NIE-IS

Lab policy P218 dictates the Cyber Security Access Controls for all computer systems within the laboratory. Paragraph 3.3.4 of P 218 states: "all local privileged accounts are denied direct access from the network." This restriction presents a problem for some Enterprise Business Systems (EBS) that require certain limited privileged accounts to traverse the network using SSH key trust relationships to automatically login to systems without a password. A security exception is in place to allow specific business procedures to use privileged accounts to communicate across the network, thereby allowing mission critical business processes, such as payroll, to operate.

The P218 exception that allows for privileged access across the network necessitates that auditing take place to ensure that only the minimum required access is permitted on these systems. Prior to the creation of the Trust Analyzer (TA), this auditing required manual analysis of the individual components within the NIE Datacenter network. The creation of the TA allows for not only the automation of this auditing function, but it also tests the functionality of the SSH keys to ensure system stability and functionality. Furthermore, the TA provides defense in depth by performing a system wide audit of all user accounts within EBS, storing the information in a central database that is easily accessed by a systems administrators or auditors with the proper credentials.

File Security Applications for Widespread Implementation on 32 and 64-bit Windows OS

Dennis Trujillo — New Mexico State University

Mentor: Jeff Dunning

Group: NIE-IS

We discuss the porting and development of applications for file security on various ports of Windows running on standard 64 (Windows Server 2008 R2) and 32-bit CPU (Windows Server 2003). Of particular focus is file security and enforcement of set privileges for shared files allowing for mitigation of security issues related to file permission parameters. Discussion relative to manipulation of Windows applicable *Discretionary Access Control* (DACL), *Access Control List* (ACL), and *Access Control Entry* (ACE) fields via memory mapping and simple string parsing procedures in C++ are referenced. Considerations of local and global (LANL-wide) security procedures are included in allusion to application of developed software. Likewise development of similar security applications is discussed.

Understanding Natural Language using Semantic Networks

Emmanuel Claudio Valentín — University of Turabo, Puerto Rico

Mentor: Jorge H. Roman

Group: HPC-1

A big problem for machines is to understand natural language. Natural Language Processing is the study of mathematical and computational modeling of various aspects of language and development of a wide range of systems. These include spoken and written languages systems. The computers don't understand human natural language. Because of that fact we want to create a tool to help the computer understand written natural language. If a computer understands natural language would be easier to process a lot of work that right now is made by humans. As the technology advances paper is becoming obsolete and the digital era is advancing. With the digital advance we obtain a lot of unstructured data. To resolve this problem we created a tool that reads text files and creates relations between the words, phrases and sentences into a database. To store relations we use Resource Descriptor Framework (RDF) triples database. RDF triples map naturally into graphs, therefore we use Graph Exploration software to display the relations. Displaying the graph facilitates the verification process to make sure the relations are created correctly. Our algorithm learns that every word has a relation to a phrase, and every phrase is related to a sentence and vice versa. This approach tells us that the more relations the software learns, the more it gives the tool a better understanding of natural language automatically. This gives us an advantage because a human takes a lot of time to process this kind of work and computers don't.

Single Hardening Tool for MacOS, Solaris, and Linux: STONIX

Derek Walker — North Carolina A & T State University

Mentor: David A. Kennel

Group: DCS-1

The DOE (Department of Energy) is among the most sought out targets of hackers around the world. To keep our data at Los Alamos National Lab (LANL) safe and secure, we must make sure the systems that hold that data are safe and secure. Manually configuring thousands of computers, like those at the lab, would be very time consuming and error prone. Automating this process increases productivity and prevents errors.

LANL uses two software applications to configure UNIX and UNIX like operating systems (OS) known as STOM (Security tool On Mac) and STOR (Security Tool On RHEL(Redhat Enterprise Linux)). Currently under development is an application known as STONIX(Software tool on *NIX) which seeks to combine the previous two applications into one platform independent hardening tool. This will allow MacOS and RHEL OS's to be hardened with a single tool and allow support to be extended to Solaris and other Linux platforms such as Mint, Debian, and OpenSUSE.

One system can take many hours to manually configure; STONIX will reduce configuration time allowing imaging technicians or other employees to configure multiple systems at once. STONIX is written in Python using object oriented programming. The design uses several well- known design patterns such as the Factory and Model-View-Controller patterns. Creating one application to do configuration for multiple OS's will also create a more effective maintenance stage of the software development life cycle by removing redundant skeleton framework and by allowing revisions to the software to occur more quickly and easily since there will now only be one application.

Server Maintenance

Michael Webber — North Carolina Agricultural & Technical State University

Mentor: Carlos Montano

Group: DCS-3

The problem I found when doing server checks monthly is trying to remember the servers we maintain and checking each one of them individually. I find these methods to take quite a long time and this time could be used elsewhere. I figured that there has to be some way to run a script that hits all the servers one owns and generate some type of report. I started looking into PowerShell, something IT professions use to do a lot of their system administration work and noticed there are many techniques one can use to find out information on servers without actually logging into each server. By writing code in PowerShell and using it to find data from the WMI (Windows Management Instrumentation) service, I'm able to display activities like, disk space usage, total Ram being used, event logs, processes, and other useful troubleshooting information. To make things simple, I created a GUI (Graphical User Interface) which is setup to use a different combination of options a user might want to display. The display comes in a report, which is sent to your email using SMTP (Simple Mail Transfer Protocol). I believe this could be a useful tool that can change how we check and maintain our servers.

Data-parallel halo finder operator in PISTON

Wathsala Nayomi Widanagamaachchi — University of Utah

Mentor: Chris Sewell

Group: CCS 7

PISTON is a portable framework which supports the development of visualization and analysis operators using a platform-independent, data-parallel programming model. Operators such as isosurface, cut-surface and threshold have been implemented in this framework, with the exact same operator code achieving good parallel performance on different architectures. An important analysis operator in cosmology is the halo finder. A halo is a cluster of particles and is considered a common feature of interest found in cosmology data. As the number of cosmological simulations carried out in the recent past has increased, the resultant data of these simulations and the required analysis tasks have increased as well. As a consequence, there is a need to develop scalable and efficient tools to carry out the needed analysis. Therefore, we are currently implementing a halo finder operator using PISTON. Researchers have developed a wide variety of techniques to identify halos in raw particle data. The most basic algorithm is the friend-of-friends (FOF) halo finder, where the particles are clustered based on two parameters: linking length and halo size. In a FOF halo finder, all particles which lie within the linking length are considered as one halo and the halos are filtered based on the halo size parameter. A naive implementation of a FOF halo finder compares each and every particle pair, requiring $O(n^2)$ operations. Our data-parallel halo finder operator uses a balanced kd-tree to reduce this number of operations in the average case, and implements the algorithm using only the data-parallel primitives in order to achieve portability and performance.

An Analysis of Source Tilting and Sub-cell Opacity Sampling for IMC

Ryan T. Wollaeger — College of Engineering/University of Wisconsin, Madison

Mentor: Todd J. Urbatsch

Group: CCS-2

Implicit Monte Carlo (IMC) is a stochastic method for solving the radiative transfer equations for multiphysics application with the material in local thermodynamic equilibrium. The IMC method employs a fictitious scattering term that is computed from an implicit discretization of the material temperature equation. Unfortunately, the original histogram representation of the temperature and opacity with respect to the spatial domain leads to nonphysically fast propagation of radiation waves through optically thick material. In the past, heuristic source tilting schemes have been used to mitigate the numerical teleportation error of the radiation particles in IMC that cause this overly rapid radiation wave propagation. While improving the material temperature profile throughout the time duration, these tilting schemes alone do not generally alleviate the teleportation error to suitable levels. Another means of potentially reducing teleportation error in IMC is implementing continuous sub-cell opacities based on sub-cell temperature profiles. We present here an analysis of source tilting and continuous sub-cell opacity sampling applied to various discretizations of the temperature equation. Through this analysis, we demonstrate that applying both heuristics does not necessarily yield more accurate results if the discretization of the material equation is inconsistent with the Monte Carlo sub-cell transport.